

Additional Table 7. Kappa scores measuring levels of agreement between the two reviewers.

<i>Nr.</i>	<i>Study*</i>	<i>Reporting</i>		<i>External validity</i>		<i>Internal validity - bias</i>		<i>Internal validity – confounding (selection bias)</i>		<i>Power</i>	
		<i>Reviewer #1</i>	<i>Reviewer #2</i>	<i>Reviewer #1</i>	<i>Reviewer #2</i>	<i>Reviewer #1</i>	<i>Reviewer #2</i>	<i>Reviewer #1</i>	<i>Reviewer #2</i>	<i>Reviewer #1</i>	<i>Reviewer #2</i>
1	Abdelnaby and Nassar 2010	8	8	0	0	5	5	4	4	4	4
2	Altuğ et al. 1989	4	4	2	2	5	5	2	2	4	4
3	Barrett et al. 2010	7	7	0	0	5	5	2	2	1.5	1.5
4	Gökalp and Kurt 2005	6	7	0	0	5	5	2	2	2	2
5	Tuncer et al. 2009	9	9	0	0	4	4	3	3	4	4
<i>Overall estimate</i>		0.750 (SE=2.100)		1 (SE=0.000)		1 (SE=0.000)		1 (SE=0.000)		1 (SE=0.000)*	

*Levels of agreement between in assessing quality scores of the included articles according to the Downs and Black (1998) scale.

**Authors in alphabetical order.

SE, Standard error